AFFTC-PA-11029

# Developing Statistically Defensible Propulsion System Test Techniques

David Kidman, Craig Stevens
Christopher Moulder, Dr. William Kitto,
Dr. James Brownlow, and Todd Remund

AIR FORCE FLIGHT TEST CENTER
EDWARDS AFB, CA

A
F
F
T
C

Approved for public release A: distribution is unlimited.

AIR FORCE FLIGHT TEST CENTER
EDWARDS AIR FORCE BASE, CALIFORNIA
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

**1. REPORT DATE** *(DD-MM-YYYY)*
15-04-2011

**2. REPORT TYPE**
Technical Presentation

**3. DATES COVERED** *(From - To)*
N/A

**4. TITLE AND SUBTITLE**
DEVELOPING STATISTICALLY DEFENSIBLE PROPULSION SYSTEM TEST TECHNIQUES

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
Kidman, David S., Propulsion Engineer
Moulder, Christopher J., Propulsion Engineer
Stevens, Craig A., Propulsion Engineer
Kitto, Dr. William, Statistician
Brownlow, Dr. James D., Statistician
Remund, Todd G., Statistician

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) AND ADDRESS(ES)**
773TS/ENFP
412 TW
307 E. Popson Ave.
Edwards AFB, CA 93524

**8. PERFORMING ORGANIZATION REPORT NUMBER**
AFFTC-PA-11029

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
773TS/ENFP
412 TW
307 E. Popson Ave.
Edwards AFB, CA 93524

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)** N/A

**12. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for public release A: distribution is unlimited.

**13. SUPPLEMENTARY NOTES**
CA: Air Force Flight Test Center Edwards AFB CA CC: 012100

**14. ABSTRACT**

Acquisition of military hardware typically proceeds from design, development, production, and finally to operational use and support. Prior to the full-rate production, both developmental and operational test and evaluation (DT&E and OT&E respectively) must occur to ensure that the system meets military requirements. The United States Air Force (USAF) is continually looking for ways to improve its test and evaluation techniques. Since 1997, Air Combat Command (ACC) has been successfully using Design of Experiments (DOE) to construct and analyze operational test efforts. This presentation highlights recent efforts to pursue statistically defensible test techniques to aid developmental test efforts.

Defensible testing is a statistical approach similar to DOE but emphasizes the need for better test planning by:

- insistence on understanding the system under test
- requiring clear and achievable test objectives
- ensuring system performance is measurable
- requiring that instrumentation accuracy and uncertainty propagation are well understood
- and requiring confidence, power, and performance thresholds

**15. SUBJECT TERMS**
defensible testing (DOE)Design of Experiments (T&E)Test and Evaluation confidence statistical analysis statistics power sample size augmentor time-to-max linear model parallel lines

| 16. SECURITY CLASSIFICATION OF: Unclassified | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| **a. REPORT** | **b. ABSTRACT** | **c. THIS PAGE** | | | 412 TENG/EN (Tech Pubs) |
| Unclassified | Unclassified | Unclassified | None | 22 | **19b. TELEPHONE NUMBER** *(include area code)* 661-277-8615 |

**Standard Form 298 (Rev. 8-98)**
**Prescribed by ANSI Std. Z39.18**

# Developing Statistically Defensible Propulsion Test Techniques

**David Kidman, Craig Stevens, and Christopher Moulder**
Propulsion Integration Flight
307 East Popson, Bldg 1400
Edwards AFB CA, 93524
Phone: (661)277-2429
david.kidman@edwards.af.mil

**Dr William Kitto , Dr James Brownlow, and Todd Remund**
Statistical Consultants Flight
306 East Popson, Bldg 1440
Edwards AFB CA, 93524
Phone: (661)277-4684
william.kitto@edwards.af.mil

Approved for public release; distribution is unlimited.
AFFTC-PA No.: 11029

**Invited Oral Presentation: 47th AIAA/ASME/SAE/ASEE Joint Propulsion Conference**
**31 July-4 August, San Diego California**

# Overview

- US Air Force Policy and Background
- Defensible Test Approach
- Case Study: Digital Engine Control Logic Upgrade
  - Test Item Description
  - Test Objective
  - Historical Approach
  - New Defensible Approach
- Challenges
- Conclusions

# Bottom Line Up Front

- Test is a *science…*not an art
  - Talented engineers but limited knowledge in test design
  - Determining statistical confidence and power allows mathematically defensible conclusions
- Air Force decisions are too important to be left to professional opinion alone…our decisions should be based on *mathematical fact*

***“I contend that all experiments are designed. Some are designed by intuition and gut feel. Other experiments … according to a rigorous statistical protocol …. In either case, experiments are designed.”—Gregory Alexander***

# AF Policy

- US Air Force policy for operational testing requires use Design of Experiments (DOE) as a discipline to improve the planning, execution, analysis and reporting. Policy states:
  - "Whenever possible, operational evaluations must include a rigorous assessment of the <u>confidence level</u> of the test, the <u>power</u> of the test and some measure of how well the test <u>spans the operational envelope</u>."[1,2]
- Air Force leadership – *"Encourages use of DOE to increase developmental test rigor"*
- Updating Defense Acquisition Guidebook to apply DOE when developing test strategies

***Currently, no formal policy requires DOE use in developmental testing***

[1]DOT&E Memo, May 2009, Subject: Test and Evaluation Initiative to Apply DOE Across Entire Acquisition Development Cycle
***https://acc.dau.mil/CommunityBrowser.aspx?id=312213&lang=en-US6***

[2]DOT&E Memo24 Nov 09, Subject: Test and Evaluation Initiatives
***http://www.dote.osd.mil/about/TE_Initiatives.pdf***

# AFFTC's Defensible History

- AFFTC Technical Advisor started effort in 2005
  - Early training from 46th Test Wing (Operational Test – OT) at Eglin AFB
  - AFFTC directive in 2007 (…implement Scientific Methods )
- Potential BUT…
  - Over simplified operational examples, hard to argue with, but application and assumptions for developmental application unclear
- AFFTC Propulsion test history
  - Tried to implement statistical approaches (Student-t, uncertainty, etc…) to compare test results, limited success
  - Slow progress improving methods
  - Now regularly working with AFFTC statistics group and AEDC

# Defensible Test Approach

AFFTC

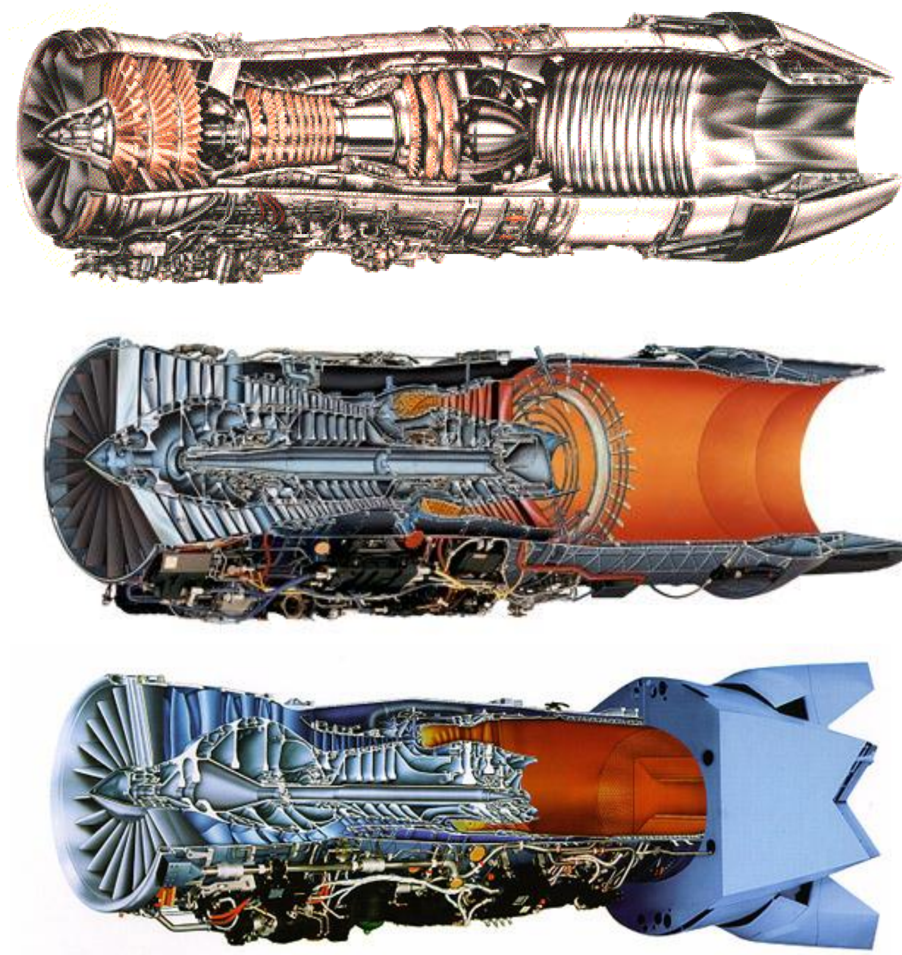
- Defensible testing is a statistical approach but also emphasizes the need for better test planning by:
  - -Understanding the system under test
  - -Defining clear and achievable test objectives
  - -Ensuring performance metrics are observable and measurable
  - -Instrumentation accuracy and uncertainty propagation is well understood.

- Statistical approach determines acceptable *Power* and *Confidence*
  - \- *Power* is the probability that the test will capture a difference between two data sets if a difference exists
  - \- *Confidence* is the probability that a prediction is correct

***Following case study highlights development of statistical approach***

# Case Study - Engine Control Upgrade

Source: http://www.edwards.af.mil/photos



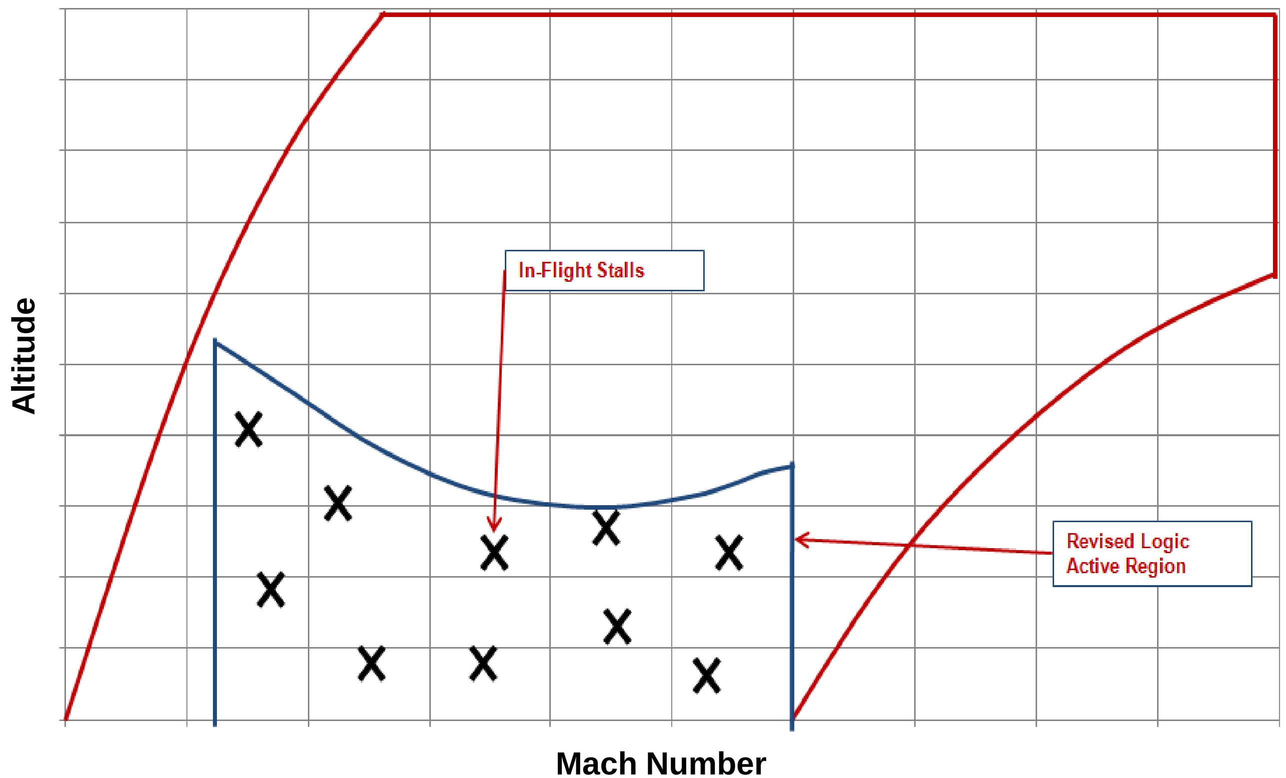
Source: http://www.aircraftenginedesign.com/

# Defensible Case Study-
# Digital Engine Control Upgrade

## Controller modified to improve stall margin in heart of A/C flight envelope

**Key logic changes:**

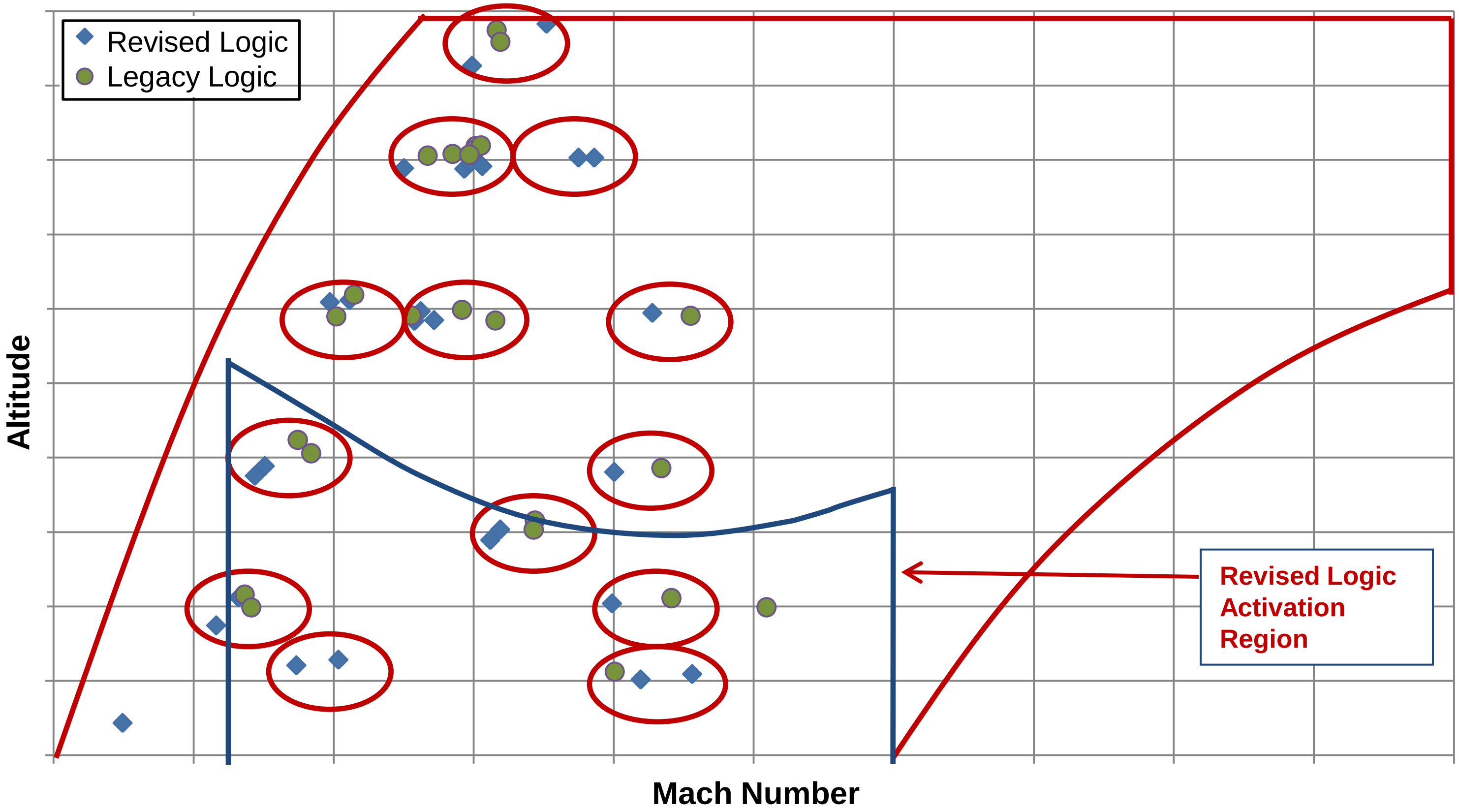
- Compressor variable vane camber was scheduled several degrees more closed
- Logic was activated and deactivated based on aircraft Mach, PT2, TT2
- Only active at high throttle settings



- **Overall Objective:** Evaluate engine stability and thrust response with revised engine control logic in comparison to the legacy engine control logic

- **Specific Objective:** Evaluate revised afterburner transient capability, specifically time-to-light and time-to-MAX and compare to legacy results

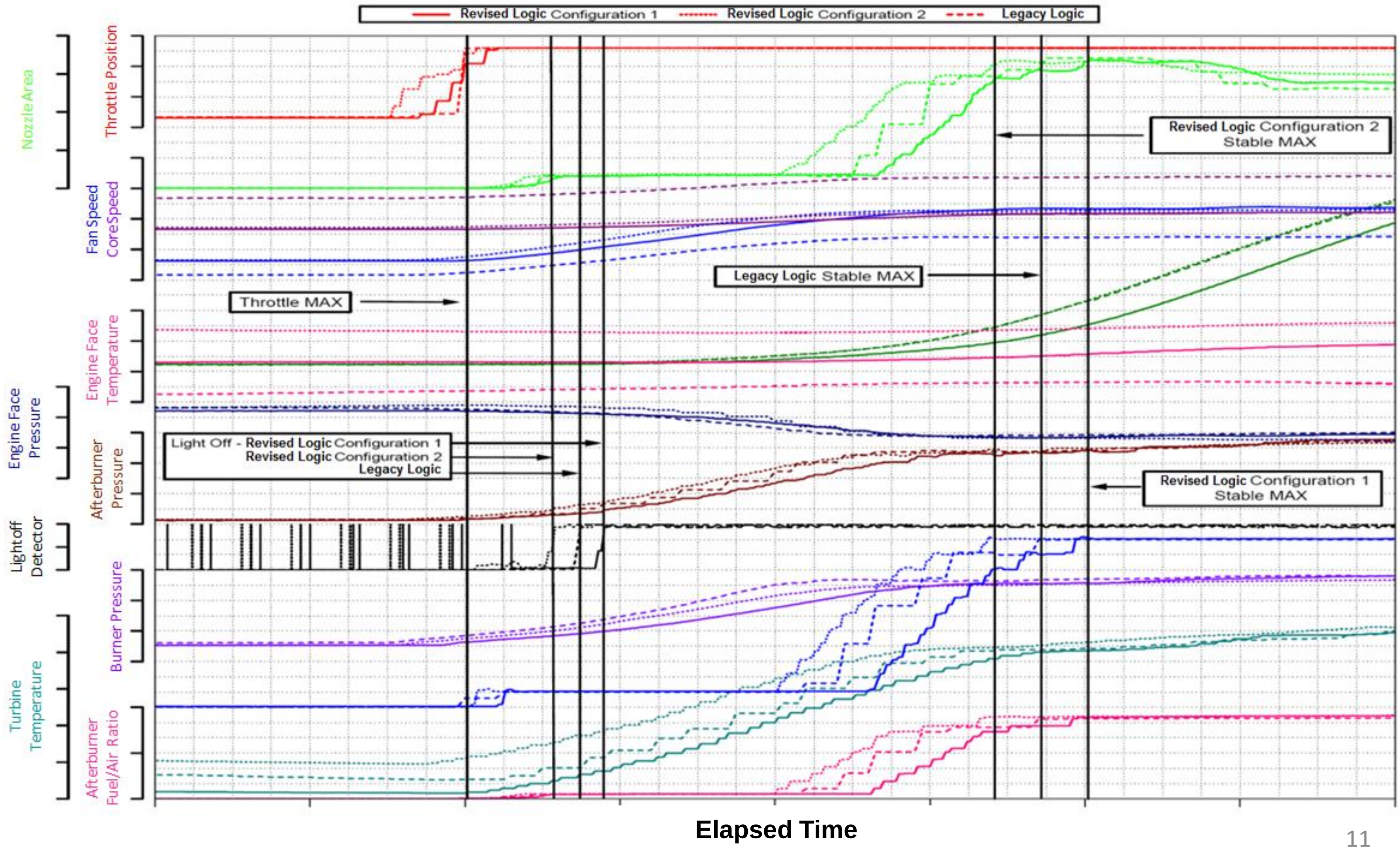
# Test Matrix




•Test matrix developed to evaluate thrust response and engine stability, focusing on the most challenging flight conditions and logic implementation areas

# Defensible Case Study- Old - Comparison Approach

## Historical Approach – Time history plot comparison IDLE-MAX



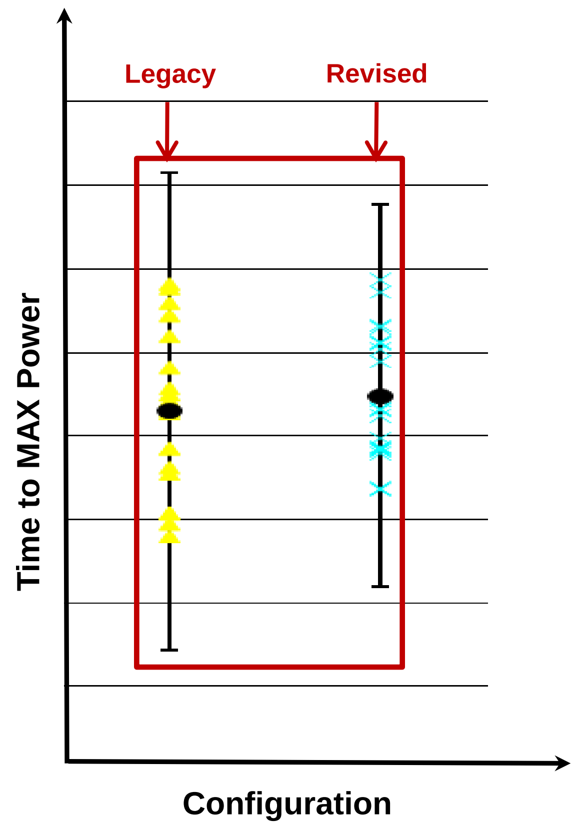
# Defensible Case Study- Old - Comparison Approach

## Historical Approach – Time to thrust comparison with 3σ error bars

- Past analysis compared thrust response times all IDLE-MAX throttle transients
    - Grouping results masked effect from logic

- Error bars didn't provide additional insight

- Final conclusion : *"In general, afterburner light-off time and time to maximum afterburner operation were comparable to the legacy logic."*



# New Defensible Techniques

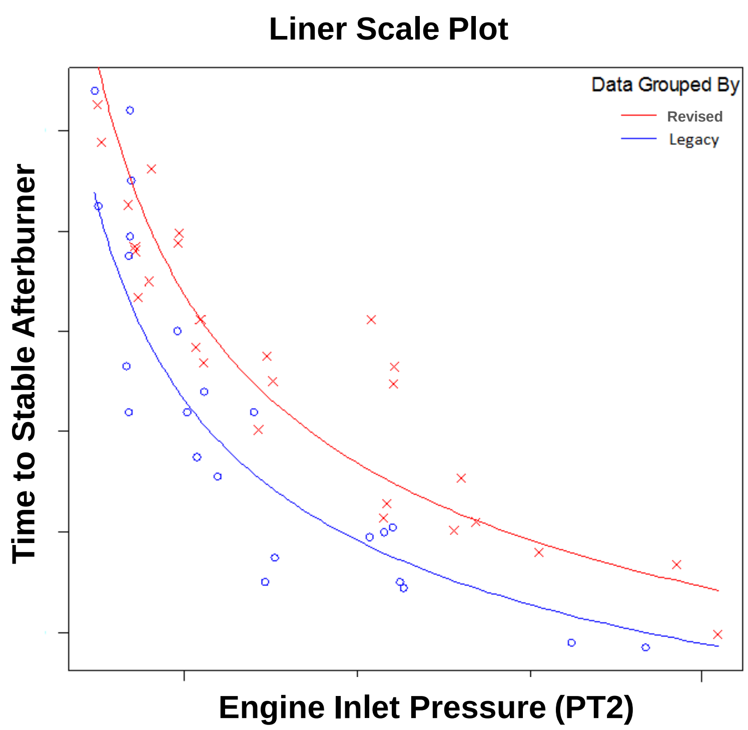
# *Defensible Case Study-*
# *New - Defensible Objective*

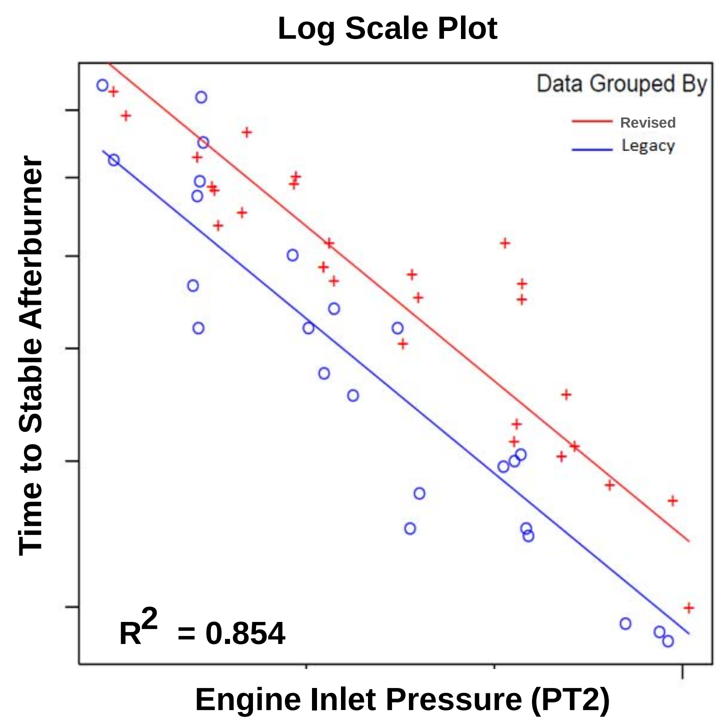
AFFTC

- **Old Specific Objective*:*** *Evaluate revised afterburner transient capability, specifically time-to-light and time-to-MAX and compare to legacy results*

- **New Specific Objective:** *Determine with statistical confidence if the revised logic thrust response has degraded in the stall avoidance flight regime as compared to the legacy logic*

# *Defensible Case Study-*
# *New Defensible Techniques*




- Experience says thrust response function of engine face total pressure (PT2)
- Power (Log-Log) transform allows model-based statistical analysis

# Defensible Case Study-
# Analysis using Inferential Model

- System was modeled such that “Logic” was an independent variable
  - If Logic = 0, then the model is

$$\ln(Time_{toAB\,Stable}) = \beta_0 + \beta_1 \ln(PT2) + \beta_2(0)$$
$$= \beta_0 + \beta_1 \ln(PT2)$$

  - If Logic = 1, then the model is

$$\ln(Time_{toAB\,Stable}) = \beta_0 + \beta_1 \ln(PT2) + \beta_2(1)$$
$$= (\beta_0 + \beta_2) + \beta_1 \ln(PT2)$$
$$= \alpha_0 + \beta_1 \ln(PT2)$$

  - If coefficient $\beta_2$ is statistically significant there is evidence that there is a difference between legacy and revised results
  - $\beta_2$ is the parallel shift difference of legacy model to the revised model.

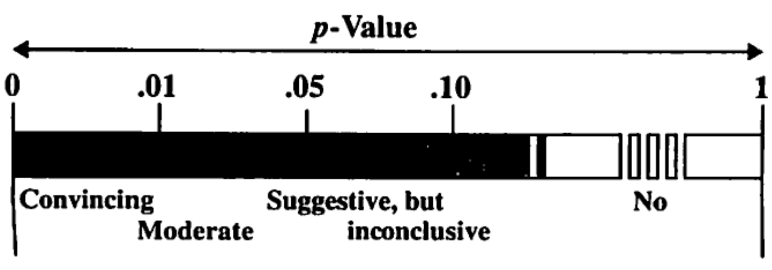
# Defensible Case Study- Parallel Lines Model

**Statistical Results: (Inside Stall Avoidance Region)**

- The t-test for the third coefficient shows a very small p-value indicating statistical confidence that a difference in thrust response exists from the legacy to revised logics
- Logarithmic units -0.2618 which translates to median time increase of 30 pct with the revised logic

| | Estimate | Std. Error | t value | p-value |
|---|---|---|---|---|
| $\beta_0$ | 2.85750 | 0.23050 | 12.397 | 2.77e-09 |
| $B_1$ | -0.54009 | 0.09095 | -5.938 | 2.72e-05 |
| $\beta_2$ | -0.26180 | 0.05946 | -4.403 | 0.000514 |

Interpreting the size of p-value | Reference: The Statistical Sleuth




Is there evidence of a difference?

# Defensible Case Study- Parallel Lines Model

**<u>Statistical Results</u>: (Outside Stall Avoidance Region)**

- The t-test for the third coefficient shows a very small p-value indicating statistical confidence that a difference in thrust response exists from the legacy to revised logics
- Logarithmic units -0.1424 which translates to median <u>time increase of 15 pct</u> with the revised logic

| | Estimate | Std. Error | t value | p-value |
|---|---|---|---|---|
| $\beta_0$ | 2.54658 | 0.07393 | 34.447 | < 2e-16 |
| $B_1$ | -0.45034 | 0.04321 | 10.422 | < 2e-16 |
| $\beta_2$ | -0.14239 | 0.03883 | -3.667 | 0.000946 |

- Outside Region should be no difference. May have been caused by engine to engine variation, aircraft installation effects, variations in flight conditions, or an actual difference caused by the engine control software

***Baseline testing with same engine and aircraft may eliminate this uncertainty***

# Defensible Case Study- A few words about… Power

**Definition:** *Power* is the probability that the test will capture a difference between two data sets if a difference exists

To capture a difference of ~ 30pct

| Power[1] | Sample Size @ Measured Std Error |
|---|---|
| 0.98 | 18 |
| 0.80 | 10 |

To capture a difference of ~ 10pct

| Power[1] | Sample Size @ Measured Std Error |
|---|---|
| 0.80 | 46 |

**Note**: 80-percent Power sufficient when failure not life threatening or causes significant financial burden

*Power is significantly affected by magnitude of difference trying to detect*

# Defensible Risks/Challenges

AFFTC

- Practical issues with sample size
  - Typical programs don't have enough time or $$$ to execute a statistically relevant test
  - Early tester involvement needed to influence test approach
- Confounding variables (life degradation, manufacturing tolerances, installation effects)
  - Test 1 engine but making fleet decisions
- Classical aspects of DOE require randomization during execution
  - Safety often requires incremental build-up for envelope expansion
- Engineers are NOT statisticians
  - It is extremely easy make erroneous conclusions with applied statistics

*Need to better understand these risks/challenges*

# Summary

- **AF moving towards policy requiring use of defensible test techniques**

- **Defensible test considerations include**
  - Test planning that addresses confidence, power, and performance threshold
  - Test objectives need to be clear and metrics measurable
  - Statistical methods vary and require quality data
  - Need to address risks/challenges identified

# Questions